

Large Scale Experiments on Fingerprint Liveness Detection

Gian Luca Marcialis¹, Luca Ghiani¹, Katja Vetter², Dirk Morgeneier², and Fabio Roli¹

¹Department of Electrical and Electronic Engineering – University of Cagliari (Italy)
{marcialis, luca.ghiani, roli}@diee.unica.it

²Crossmatch Technologies Inch.
{katja.vetter, dirk.morgeneier}@crossmatch.com

Abstract. Fingerprint liveness detection consists in extracting measurements, from a fingerprint image, allowing to distinguish between an "alive" fingerprint image, that is, an image coming from the fingertip of the claimed identity, and an artificial replica. Several algorithms have been proposed so far, but the robustness of their performance has not yet been compared when varying several environmental conditions. In this paper, we present a set of experiments investigating the performance of several feature sets designed for fingerprint liveness detection. In particular we assessed the decrease of performance when varying the pressure and the environmental illumination as well as the size of the region of interest (ROI) used for extracting such features. Experimental results on a large data set show the different dependence of some features sets on the investigated conditions.

1 Introduction

Identification of a person based on the so-called biometrics, namely physical (fingerprints, face, iris) or behavioural (gait, signature) attributes is an alternative paradigm to those relying on what he/she possesses (e.g. a card that can be lost or stolen) or remembers (e.g. a password that can be forgotten) [1]. Nowadays, more than ever, it is very important to be able to tell if an individual is authorized to perform actions like entering a facility, access privileged information or even cross a border. Therefore, biometric systems are considered to be more reliable for the recognition of a person than traditional methods.

A biometric system is a pattern recognition system that acquires biometric data from an individual, extracts a features set from the data, compares these features against those stored in a database and executes an action based on the comparison result.

Fingerprints are the most used, oldest and well-known biometric measurements [2]. Fingerprints exhibit important properties as uniqueness and permanence. They are composed of epidermic ridges and valleys flow, which smoothly varies around two or more singular points named *core* and *delta*.

Although fingerprints were often claimed difficult to be steal and reproduced, it has been recently shown that artificial replication is possible [3]. Furthermore, the related image obtained by electronic sensors can be difficult to distinguish from "alive" ones, even by visual inspection. Therefore, the development of "liveness" detection techniques is important to try to distinguish if a fingerprint image is coming from an alive person or from a replica. Liveness detection seeks additional data to verify if a biometric measure is authentic. Fingerprint liveness detection, with either hardware-based or software-based systems, is used to check if a presented fingerprint originates from a live person or an artificial finger [4]. It is based on the principle that additional information can be obtained from the data acquired by a standard verification system. This additional data can be used to verify if an image is authentic.

To detect liveness, hardware-based systems use additional sensors to gain measurements outside of the fingerprint image itself while the software-based ones use image processing algorithms to gather information directly from the collected fingerprint. These systems classify images as either live or fake [4-9].

Software-based approaches are cheapest than hardware-based, since these require additional and invasive hardware to measure the liveness directly from the fingertip of people. Instead, software-based must detect liveness from features extracted from the fingerprint images captured by the sensor. In other words, the liveness detection problem is treated as a pattern recognition problem, where a set of features must be selected in order to train an appropriate classifier.

Although several feature sets have been proposed to this aim, it is difficult to assess the state-of-the-art appropriately. Moreover, the variables to be taken into account are so much, that it is often impossible to perform an exhaustive and fair comparison among methods: for example, the sensor type, materials used for fabricating the fingerprint replicas, the environmental conditions as temperature, illumination, the ability of the attacker in pressing the replica on the sensor surface, the fingerprint region used to extract liveness features (ROI), and so on.

Therefore, in this paper we assess a fair comparison of several state-of-the-art approaches to fingerprint liveness detection on a large data set made up live and fake fingerprint images acquired by the Crossmatch sensor LSCAN Guardian USB. In particular, after analyzing the baseline performance of such algorithms, we focus on three environmental conditions: illumination, pressure and selected ROI. In all cases, we measure the effect on the system performance and point out several countermeasure in order to improve the system robustness.

This paper is organized as follows. Section 2 briefly describes the investigated algorithms. Section 3 describes data set, protocol and experiments performed al results and performance obtained. Section 4 concludes the paper.

2 Investigated algorithms and open issues

In this paper, we reported experimental results on several state-of-the-art fingerprint liveness detection algorithms. We briefly describe them in the following. Further details can be found in the related references.

Local Binary Patterns (LBP) [5]: local binary patterns were first employed for two-dimensional textures analysis and excellent results were obtained due to their invariance with respect to grey level, orientation and rotation. It extracts certain uniform patterns corresponding to micro-features in the image. The histogram of these uniform patterns occurrence is capable of characterize the image as it combines structural (it identify structures like lines and borders) and statistical (micro-structures distribution) approaches. According to [5], a 54-sized feature vector has been obtained.

Power spectrum [6]: Coli *et al.* analyzed fingerprints images in terms of high frequency information loss. In the artificial fingerprint creation, the ridge-valley periodicity is not altered by the reproduction process but some micro-characteristics are less defined. Consequently, high frequency details can be removed or strongly reduced. It is possible to analyze these details by computing the image Fourier transform modulus also called “power spectrum”. We selected twenty sub-bands on the power spectrum, so obtaining a 20-sized feature vector.

Wavelet energy signature and Gray-Level Co-occurrence Matrix (GLCM) [7]: The gray-level co-occurrence matrix (GLCM) takes account of how often a pixel with gray-level (grayscale intensity) value i is adjacent to a pixel with the value j . Actually, the element (i, j, d, θ) represents the probability that a couple of pixels x, y at distance d and orientation θ have gray levels i and j respectively. We considered a distance $d = 1$, so the GLCM matrix is related of local characteristic of the image, and four orthogonal directions for θ , as done in [7]. Therefore we computed four matrices $C_\theta(i, j)$, and, for each of them, a group of ten features, so obtaining a 40-sized feature vector.

Wavelet 2D [8]: wavelet decomposition of an image lead to the creation of four sub-bands: the approximation sub-band containing global low frequency information, and three detail sub-bands containing high frequency information. The image is decomposed in four levels, three sub-bands for each one, and three different wavelet filters (Haar, Daubechies (db4) and Biorthogonal (bior2.2)), so obtaining a 70-sized feature vector.

Curvelet [9] decomposition is very efficient for representing edges and other singularities along fingerprint ridges due to his high directional sensitivity and his high anisotropy. We consider two different sets of features, also called “signatures” in [9]:

- Curvelet energy signature: the energies of the 18 sub-bands are measured by computing means and variances of curvelet coefficients.
- Curvelet co-occurrence signature: for each of the 18 sub-bands, the GLCM (Gray Level Co-occurrence Matrix) is calculated together with 10 corresponding features.

In order to test the accuracy of these algorithms, Refs. [5-9] use appropriate data sets, but different, one each others, with respect to the size, the materials used for replicating fingerprints. Classifiers used are different too. Therefore, methods cannot be compared by simply considering results reported in those papers.

Moreover, no experimental investigation has been done on the environmental conditions affecting the liveness detection performance, especially if these algorithms must be intergrated in real fingerprint verification systems [8]. From this point of view, we can consider characteristics “intrinsic” to the feature set chosen, like the

location of the region of interest selected for feature extraction, and “intrinsic” to the sensor adopted that, like the pressure of the attacker on the fingerprint sensor surface, and the environmental illumination. These points may impact on the quality of the liveness feature set extracted, thus being crucial to analyze the “robustness” of such feature sets against attacks based on fake fingerprints. In fact, if the system can be less robust where environmental illumination changes, a person can take advantage of this, by choosing the best moment to attack the system or modifying the environmental light. The same holds for the pressure. Finally, if the features are sensitive to the ROI position, a wrong ROI extraction could lead to misclassification errors. However, it is unknown at which extent they are important, and, eventually, which countermeasures can be adopted to reduce their impact. This is the scope of the present paper, where these characteristics are analyzed by experiments, and some preliminary observations are drawn from the obtained results.

3 Experimental results

3.1 Data sets and experimental protocol

We used a four data sets for our investigations:

- 1) D.1 – this data set is made up of 1816 live fingerprints and 1624 fake fingerprints made up of commonly used materials, uniformly distributed along replicas: silicone, gelatin, wood glue and latex. Molds are made up of plastiline-like material which allows to replicate the 2d contour of the fingertip. Fingerprint images have been acquired by the Crossmatch LSCAN Guardian USB electronic sensor. The data set has been subdivided in two parts, namely, training set and test set, according to the protocol adopted in the recent Second edition of Fingerprint Liveness Detection Competition (LivDet2011) [10]. A multi layer perceptron (MLP) has been trained on the first part of data, so obtaining the *baseline fingerprint liveness detector*. The MLP output is interpreted, as usual, as liveness detection score in the range [0,1]. Features sets are extracted from ROIs located on the core of fingerprint images (the core is centre of the fingerprint image according to [1]), as shown in Fig. 1(a). Such ROIs are quadrangular regions. Two sizes has been used: 80x80 pixels and 160x160 pixels.
- 2) D.2 – this data set has been built to test the variability of the feature set performance when ROIs are not correctly located. Four different location errors are studied, as reported in Fig. 1(b). We tested the performance on the baseline system with 80x80 pels ROIs, but also the performance which can be obtained by adding to the training set also patterns extracted from wrongly located ROIs.
- 3) D.3. This data set has been built for evaluating the impact of the pressure of the fake fingerprint on the sensor surface when baseline system is used with 80x80

pels for ROIs. In order to generate novel images, we put a increasing weight from 500 g to 4000 g on the fake fingerprint, thus simulating the different pressure. Obtained data set is thus made of 500 fake fingerprint frames (with increasing weight over frames) per three different types of silicon-like materials, gelatin and latex for replicating fingerprints. On overall, 2,100 test images have been used (not overlapped with the training/test set used for the baseline system). Effect is studied by evaluating the variation and correlation degree of the liveness detection score with the related weight on the fake fingerprint.

- 4) D.4. This data set has been built for evaluating the impact of the environmental illumination on the features sets related to fake fingerprints. It has been organized as follows: 103 fake fingerprint images simulate device initialization in dark room without any enviromental illumination (condition 1); 103 fake fingerprint images simulate device initialization with directed light (condition 2). Influence of environmental illumination has been tested by evaluating average and standard deviation of the liveness score for conditions 1-2.

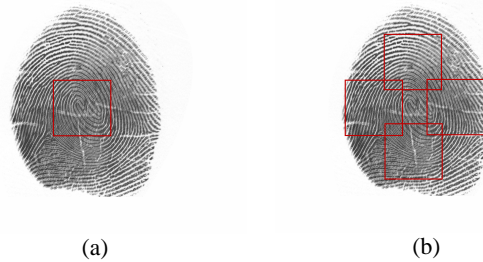


Figure 1. ROI positions. (a) Baseline ROI. (b) Wrong locations: up, down, left, right.

3.2 Baseline results

Figs. 3(a-b) show the ROC curves of the baseline system according to D.1 data set.

It can be seen that LBP feature set leads to the best performance, whilst the feature set Power spectrum one leads to the worst one. Experiments show that the error slightly depend on the size of the ROI. Doubling the size of the ROI, that is, from 80x80 to 160x160 pixels, the error decreases of about 3% on average. It is worth noting that, using a ROI of 160x160 pixels, a fraction of background could be present during feature extraction, but this does not seem relevant on the basis of reported results.

Moreover, the rank of investigated feature sets, from the best one to the worst one, is: LBP, Wavelet2D, Curvelet Energy, Curvelet Glcm, GLCM, PS. It is independent on the ROI size. Therefore, all feature sets are sensitive to the ROI size, in a very similar manner.

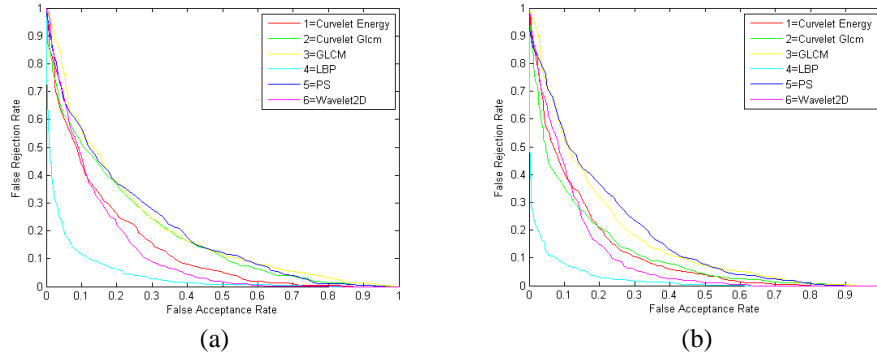


Figure 3. ROC curves showing the expected performance on the D.1 test set. (a) ROI size: 80x80 pixel. (b) ROI size: 160x160 pixel.

3.3 Performance on wrongly located ROIs

Table 1 summarizes results as follows. For each feature set (first column), EER is reported when training the classifier on centered images (second column), that is, on the baseline system. This is used as reference result. Third column reports the same classifier when tested on different ROI positions.

It is possible to see that a wrong ROI position weakly affects the system performance (second-third column of Table 1). In all cases, a loss of performance is about 1%. In particular, LBP and Wavelet appear as the preferred feature sets.

In order to recover this performance difference, one could think that training set should be “empowered” with the addition of wrongly located ROIs. Therefore, we re-trained the baseline classifier with such novel information, and test the performance of the same test sets. Results are reported in Table 2. It is easy to see that recovering above performance variation by adding to the training set bad centered images is not possible.

These results allow to observe that ROI location does not appear a very crucial point for all feature sets, so estimation errors of the ROI do not impact on the final system performance, independently of features sets adopted among the ones investigated here.

Table 1. EER for all feature sets and different ROI positions.

| Feature set | Baseline test | Wrongly located ROIs test |
|-----------------|---------------|---------------------------|
| LBP | 10.90 | 12.45 |
| GLCM | 27.32 | 28.03 |
| Wavelet2D | 20.94 | 22.07 |
| Curvelet Energy | 23.34 | 24.56 |
| Curvelet GLCM | 27.32 | 27.99 |
| Power Spectrum | 28.84 | 28.59 |

Table 2. EER for all features sets on the classifier “empowered” with patterns extracted from wrongly located ROIs.

| Feature set | Empowered classifier | Empowered classifier |
|-----------------|---------------------------|----------------------|
| | Wrongly located ROIs test | Baseline test |
| LBP | 12.66 | 11.84 |
| GLCM | 26.07 | 25.68 |
| Wavelet2D | 21.36 | 20.21 |
| Curvelet Energy | 24.15 | 23.74 |
| Curvelet GLCM | 27.31 | 26.66 |
| Power Spectrum | 29.24 | 29.76 |

3.4 Effect of pressure variation

We summarise basic results in Table 3, where we report, for each material and each feature set, the sign of the correlation between the system output (1 – liveness score) and the applied pressure. If this correlation is more than 0.5 we write ‘+’ on the related cell in Table 3; if it is less than -0.5, ‘-’ is written; if correlation is averagely “low”, that is, between -0.5 and 0.5, we indicate this fact by ‘*’.

The most promising feature set is still LBP, because is positive in almost every case. This means that, if a fingerprint is spoof, the more the pressure, the more this evidence. The same holds for the live class. This can be also desired in optical sensors as the one adopted for these tests, because the more the pressure, the more the sharpness of related image. Worth noting, all feature sets are positively correlated with pressure when live fingerprints are submitted. This is good, but the decrease of the posterior probability for the fake class may lead to an increase of false acceptance rate (see for example Wave and PS columns).

Table 3. Positive (+), negative (-) or no correlation (*) between system output (posterior probability of the correct class) and the applied pressure

| | LBP | GLCM | Wavelet2D | Curv. Energy | Curv. GLCM | P.S. |
|------------|-----|------|-----------|--------------|------------|------|
| Silicone 1 | + | + | - | * | * | - |
| Silicone 2 | * | - | - | - | * | - |
| Silicone 3 | + | + | - | + | - | * |
| Gelatine | + | - | - | + | * | - |
| Latex | + | - | - | + | * | - |

We reported in Table 4 if the system outputs, for each feature set, fall into the realted average range of spoof and fingers on D.1 Test. Symbols in Table 4 can be interpreted as follows:

“+”: object is always in range from 500g-4000g;

“*”: object is not in range for each weight, but there is a visible tendency (correlation exists);

“-”: object is not in range for each weight, but there is not a visible tendency (no correlation exists).

Reported results largely confirm observations from Table 3. LBP, Curvelet GLCM and Curvelet Energy related outputs fall in the standard range in the most of cases, thus they cannot be considered as preferred feature sets.

Table 4. Positive (+), negative (-) or no correlation (*) between system output (posterior probability of the correct class) and the applied pressure when considering only standard output range (D1.1. Test) and related outputs.

| | LBP | GLCM | Wavelet2D | Curv. Energy | Curv. GLCM | P.S. |
|------------|-----|------|-----------|--------------|------------|------|
| Silicone 1 | + | + | * | - | + | + |
| Silicone 2 | + | * | * | + | + | * |
| Silicone 3 | * | * | * | + | + | - |
| Gelatine | + | * | * | + | - | + |
| Latex | * | - | * | + | + | + |

3.5 Effect of environmental illumination

Results are shown in Table 5. Conditions 1-2 are the ones explained in Section 3.1. In all cases, similar output values, that is, liveness score, are obtained, thus illumination does not appear as a relevant environmental condition for the system output.

On the basis of available data set and reported experiment, LBP and Wavelet feature sets appear as the preferred ones.

Table 5. Posterior probabilities of spoof samples by varying the illumination conditions, and related standard deviation (in brackets).

| | Liveness score average and standard deviation | | |
|--------------|---|---------------|---------------|
| | Baseline | Condition (1) | Condition (2) |
| LBP | 0.960(0.019) | 0.962(0.066) | 0.969(0.028) |
| GLCM | 0.706(0.008) | 0.660(0.066) | 0.666(0.081) |
| Wavelet2D | 0.941(0.020) | 0.934(0.022) | 0.932(0.019) |
| Curv. Energy | 0.788(0.024) | 0.747(0.043) | 0.762(0.043) |
| Curv. GLCM | 0.684(0.064) | 0.630(0.097) | 0.659(0.084) |
| PS | 0.765(0.021) | 0.805(0.028) | 0.813(0.019) |

4 Conclusions

In this preliminary set of large scale experiments on fingerprint liveness detection, we focused on the baseline system performance of several state-of-the-art feature sets, with respect to some of variability elements. In particular, we have studied an intrinsic characteristic of the feature sets, namely, the choice of the ROI (size and location), and two external characteristics of the fingerprint sensor, that is, the pressure of the

fake fingerprint on the sensor surface, and the environmental illumination, which may impact on the captured image, and, thus on the feature set extracted.

On the basis of reported experiments, we noticed that, in almost all cases, the large is ROI size, the better is the system performance, but the location of the ROI is not relevant. It has also been obtained that the most of features sets are sensitive to the pressure, thus improving or worsening the liveness detection result depending on how much an individual tend to press the fake fingerprint (see in particular the LBP and Curvelet Energy cases). Finally, environmental illumination is not crucial, since the system output is substantially stable independently on rough changes of the light intensity.

These results point out that performance is not yet acceptable for their integration in standard fingerprint verification algorithms, but also that they have some invariant characteristics to some settings and environmental conditions which make them worthy of further theoretical and experimental investigations.

Acknowledgments

This work has been supported by the sponsored research agreement between University of Cagliari and Crossmatch Technologies Inc.

References

1. A.K. Jain, P. Flynn, A. Ross, *Handbook of Biometrics*, Springer, 2007, ISBN 9780387710402.
2. D. Maltoni, D. Maio, A.K. Jain, S. Prabhakar, *Handbook of Fingerprint Recognition*, Springer, New York 2003, ISBN 0387954317.
3. T. Matsumoto, H. Matsumoto, K. Yamada, H. Hoshino, Impact of artificial 'gummy' fingers on fingerprint systems, *Proceedings of SPIE*, vol. 4677, 2002.
4. P. Coli, G.L. Marcialis, F. Roli, Vitality Detection from Fingerprint Images: A Critical Survey, *IEEE/IAPR 2nd International Conference on Biometrics ICB 2007*, D.O.I. 10.1007/978-3-540-74549-5_76.
5. S.B. Nikam, and S. Aggarwal, Local binary pattern and wavelet-based spoof fingerprint detection, *Int. J. of Biometrics*, 1 (2) 141-159, 2008.
6. P. Coli, G.L. Marcialis, F. Roli, Power spectrum-based fingerprint vitality detection, *IEEE Int. Work. on Automatic Identification Advanced Technologies AutoID 2007*, pp.169-173.
7. S.B. Nikam, and S. Aggarwal, Wavelet energy signature and GLCM features-based fingerprint anti-spoofing, *IEEE Int. Conf. On Wavelet Analysis and Pattern Recognition*, 2008. DOI: 10.1109/ICWAPR.2008.4635872.
8. A. Abyanka, and S. Schuckers, Integrating a wavelet based perspiration liveness check with fingerprint recognition, *Pattern Recognition*, 42 (2009) 452 – 464.
9. S.B. Nikam, and S. Agarwal, Fingerprint Liveness Detection Using Curvelet Energy and Co-occurrence Signatures, *Fifth International Conference on Computer Graphics, Imaging and Visualization 2008 IEEE*, D.O.I. 10.1109/CGIV.2008.9.
10. D. Yambay, et al., LivDet 2011 - Fingerprint Liveness Detection Competition 2011, 5th IAPR/IEEE Int. Conf. on Biometrics (ICB 2012), in press.